EuJAP | VOL. 6 | No. 1 | 2010
ORIGINAL SCIENTIFIC PAPER
UDK 1 Jaspers, K.
159.97
616.89:1

ARE PSYCHIATRIC KINDS 'REAL'?

HELEN BEEBEE and NIGEL SABBARTON-LEARY
University of Birmingham

ABSTRACT

The paper considers whether psychiatric kinds can be natural kinds and concludes that they can. This depends, however, on a particular conception of 'natural kind'. We briefly describe and reject two standard accounts – what we call the 'stipulative account' (according to which apparently *a priori* criteria, such as the possession of intrinsic essences, are laid down for natural kindhood) and the 'Kripkean account' (according to which the natural kinds are just those kinds that obey Kripkean semantics). We then rehearse a more permissive account: Richard Boyd's 'homeostatic property cluster' (HPC) account. We argue that psychiatric kinds can in principle count as natural kinds on the HPC account. Moreover, specific psychiatric kinds (Tourette's, schizophrenia, etc.) can be natural kinds even if the category *psychiatric disorder* is not itself a natural kind.

Keywords: natural kind, psychiatric kind, homeostatic property cluster

1. Introduction

Since Thomas Szasz's charge that the concept of mental illness is merely a 'convenient myth' (1961, 113), psychiatrists have been concerned with the issue of whether mental categories typically thought to be disorders (e.g. autism, Tourette's, schizophrenia and depression) are 'real' or 'objective'. To use a standard metaphor, do such categories 'carve nature at the joints'?

One way to frame the debate is in terms of natural kinds, where natural kinds are understood broadly as independently existing divisions in nature (the 'joints' at which we aim to carve the natural world). Paradigmatic scientific research programmes, such as physics, chemistry and biology, all appear to classify the objects of their investigation into natural kinds. Physics classifies subatomic particles into distinct kinds (photon, electron, muon), chemistry carves substances into discrete types via their molecular structure and composition, and biology classifies individual organisms according to their species, genus and family. The classification of objects into natural kinds, then, appears to be one of the marks of a *bona fide* scientific research programme, since natural kinds figure

fundamentally in classification, prediction and explanation.

The alleged problem for the kinds found in psychiatry is that they are merely conventional rather than 'objective'. Thus Derek Bolton says: "the status of the norms invoked in demarcating mental disorders ... has been controversial ... are they a matter of objective, medical fact – or [are] they really social?" (Bolton 2008, 4). Indeed, some have contrasted the 'objective' not only with the 'social', as Bolton does, but with arbitrariness: as Peter Zachar puts this view (which he does not endorse): "if a category can't be conceptualized as a natural kind, it is an arbitrary category" (Zachar 2000, 167). The influence of the social element of demarcation has led many authors (including Zachar) to withhold the epithet 'natural' from the kinds of psychiatry and psychology, often claiming instead that such kinds are arbitrary (McCrae 1994), mythical (Szasz 1961), and fundamentally unreal. This in turn raises a worry about the scientific credentials of psychiatry: if psychiatric kinds are not natural kinds, the status of psychiatry as a *bona fide* scientific discipline might be thought to be in question.

Our contention in this paper will be that there is no principled metaphysical reason to deny that psychiatric kinds are natural kinds. This leaves open the possibility that there are good reasons to think that some or all of the categories *currently* recognised by psychiatry are not natural kinds because we have good reasons to think that a significant amount of reclassification will occur in the future, on the basis of additional evidence, new treatments, and so on. But this is an issue we shall leave aside.

We shall proceed as follows. In §2, by way of setting the scene for the rest of the paper, we make some clarificatory remarks. In §3, we draw a rough contrast between supposedly 'natural' and 'conventional' kinds. We then rehearse and briefly criticise one standard approach to delineating the natural kinds – what we shall call the 'stipulative account', according to which we legislate (presumably on the basis of *a priori* philosophical reflection) the characteristics that natural kinds must have. In §4 we do the same for a second standard account – the 'Kripkean account' – which appeals to a broadly Kripkean story about the necessary *a posteriori* status of so-called 'theoretical identity statements'. On this view, natural kinds are distinctive in that (a) members of a natural kind share a 'real essence', and (b) it is a matter of discovery, and not stipulation, what the essence of a given natural kind is. In §5, we sketch an alternative account of natural kinds – Richard Boyd's 'homeostatic property cluster' account – which focuses on causally stable classification and hence predictive and explanatory success. We argue that, at least *prima facie,* psychiatric kinds can in principle count as natural kinds in this sense, despite their lack of shared underlying real essence.

Finally, in §6, we address the worry that the notion of a 'disorder' is socially constructed, and so, even by the lights of Boyd's account of natural kinds, psychiatric disorders cannot count as natural kinds. We argue that even if this is so, it is no bar to thinking of specific kinds of disorder (Tourette's, schizophrenia, and so on) as themselves natural kinds.

2. Clarificatory remarks

In this section, we set the scene for the ensuing discussion by making three general clarificatory remarks.

First, we note that the question, whether certain mental categories (Tourette's, schizophrenia, and so on) are natural kinds is distinct from the question whether the additional classification of those categories as 'disorders' itself demarcates a natural kind. To illustrate, take the analogy with the kind *jade*. Something is jade just if it is one or other of two chemically distinct substances, jadeite and nephrite – and it is (let us suppose) uncontroversial that these kinds of substance are each, themselves, natural kinds.[1] This just by itself, however, leaves it entirely open whether *jade* itself is a natural kind: as will become clear later on, some philosophers would deny that jade is a natural kind, while others are happy to grant that it *is* a natural kind. Similarly, we can in principle accept that (say) Tourette's and schizophrenia are natural kinds while accepting that categorisation of those conditions as *disorders* is not itself a natural-kind classification (perhaps because the notion of a disorder is socially constructed in a way that is incompatible with regarding *disorder* as a natural kind). We return to this issue in §6 below; for now, however, we simply want to put the distinction on the table. This is because our main aim in this paper is to argue that at least some psychiatric categories can in principle be natural kinds; the point of the distinction is that this can be so even if the category *disorder* is not itself a natural kind.

Second, we shall deliberately avoid characterising the debate about the status of psychiatric categories as a debate about whether such categories have 'essences'. This is a standard way of framing the debate, both in psychiatry and in biology.[2] We believe, however, that it is not a helpful way to proceed. What an 'essence' is supposed to be varies enormously across the philosophical literature; so a discussion of whether or not (say) biological kinds 'have essences' is unhelpful unless the notion of essence itself is clearly defined (as it often is not). For example, it is sometimes assumed that in order for a kind to have an 'essence', there must be some single intrinsic feature or set of features that all and only members of the kind share (often such features are assumed to be at least partly a matter of constitution, that is, what 'stuff' something is made of). On a weaker (but distinct) notion of essence, essences must be intrinsic only in the sense that they must be non-relational: what makes something a member of an essence-involving kind is purely a matter of its internal constitution. That these are two distinct notions of 'essence' can be seen from the case of jade: there is no *one* set of intrinsic (chemical) features that all samples of jade share; on the other hand, whether something is a sample of jade does nonetheless depend solely on its internal (chemical) constitution.

[1] Actually it is (or should be, at least according to the stipulative and Kripkean accounts discussed below) controversial whether jadeite and nephrite are themselves natural kinds, given that there can be a good deal of variation of chemical constitution within each kind. For example, nephrite itself has variable chemistry: anything on the spectrum between tremolite – $Ca_2Mg_5Si_8O_{22}(OH)_2$ – and ferroactinolite – $Ca_2Fe_5Si_8O_{22}(OH)_2$ – counts as nephrite, where the magnesium (Mg) in tremolite might be partially or wholly replaced by iron (Fe).

[2] For a useful summary of the role of debates about 'essences' in the debate about biological classification, see Griffiths 1999. For discussion of essences in the debate about psychiatric classification, see for example Cooper 2005, Chapter 2.

On a very different conception of 'essence', a kind has an essence simply if there is some sharply defined set of criteria – that is, necessary and sufficient conditions – for kind membership. On this notion of essence, the kind *sphere* has an essence, despite the fact that two members of the kind may have utterly different internal constitutions (a sphere can be made of gold, or iron, or …).

If (as is often done) we add the claim that the *natural* kinds are those kinds that have essences, we can see how unhelpful the appeal to essences is. Whether or not a kind will count as natural will depend in turn on our choice of what we mean by 'essence'; but how are we to decide which notion of 'essence' is the appropriate one to use? We suspect that at this point the parties to the debate will choose whichever notion of 'essence' satisfies their intuitions about which kinds are the natural kinds, so that, for example, if one holds that, intuitively, jade is not a natural kind, then one will adopt a notion of essence that requires sameness of chemical constitution; whereas if one holds that jade *is* a natural kind, one will adopt a less demanding notion of essence. This being so, appeal to the notion of essence is really a distraction: we would be better off thinking directly about which kinds are natural kinds and why this is so, rather than attempting to locate some preferred notion of 'essence' and defining 'natural kind' in terms of that preferred notion.

Third, and finally, we note that there are two very different ways in which we might approach the question of what counts as 'carving nature at its joints'. Suppose we start out with the thought that the fundamental 'joints in nature' are delivered by physics: it is physics, after all (let us suppose), that tells us what the fundamental constituents of matter are, such that everything – whether it is a molecule of hydrogen or an amoeba or a garage – is composed of those constituents. This starting-point encourages what might be thought of as a 'constitutional' conception of natural kinds: a conception of natural kinds according to which it is what something is *made* of that determines which natural kind it belongs to (if any). Thus, for example, Colin McGinn argues that 'functional' kinds – such as mental-state kinds and artefact kinds – cannot be natural kinds, because such kinds "are not possessed of an *a posteriori* physical real essence constituting them as of the kind they are" (1978, 199). Whether or not something counts as *money*, for example, depends upon the functional role it plays in the exchange of goods, and that role can be performed by indefinitely many kinds of stuff (gold, bronze, pieces of paper, and so on). Thus money lacks an '*a posteriori* physical real essence': some underlying physical nature that is shared by all and only members of the kind.[3]

By contrast, we might start from a very different place, with the thought that the natural kinds are just those classifications that are especially explanatorily and predictively successful. Take the fact that the periodic table classifies the elements in terms of atomic number (that is, number of protons in the atomic nucleus). Carbon-12 and carbon-13, for example, are both (unsurprisingly) isotopes of carbon – they both have atomic number 6 – but they have different numbers of neutrons (6 and 7 respectively). The system of classification that defines elements on the basis of atomic

[3] McGinn's argument appeals to the Kripkean account discussed in §4 below.

number is better than a system that does so on the basis of, say, number of neutrons because, by and large, sameness of atomic number goes along with sameness of chemical behaviour: the periodic table, as it actually is, is a much better way of classifying *because* it is explanatorily and predictively better. If we abstract away from explanatory and predictive success and focus just on the *constitution* of atoms, however, there are simply no grounds for regarding our existing classification as 'the' natural one. True, all carbon atoms have something in common, namely 6 protons. But then, carbon-14 and oxygen-16 also have something in common, namely 8 neutrons. If we're solely interested in the constitution of atoms, there is no reason to classify in terms of atomic number rather than number of neutrons; after all, there is nothing special about protons, just *qua* constituents of atoms, that dictates that classification in terms of number of protons is more 'natural' than classification in terms of number of neutrons.

If we take the periodic table as a paradigmatic natural-kind classification, then, it seems that the distinctive features of natural kinds that we should focus on are explanatory and predictive success, rather than sameness of underlying constitution. And – or so we shall argue – this approach to natural kindhood is much more promising than the alternative when it comes to considering the status of psychiatric kinds. After all, as we have already seen, the 'constitutional' conception of natural kinds looks likely to render *all* mental categories non-natural kinds, assuming that mental kinds are functional kinds. As we shall see, the constitutional conception also runs into trouble when it comes to biological classification. By contrast, if we adopt the non-constitutional conception, it might well turn out that psychiatric categories can in principle deliver sufficient explanatory and predictive success to count as natural kinds. And, as we have just seen, adopting this conception of natural kinds is no *ad hoc* manoeuvre designed solely to rescue psychiatric kinds from metaphysical oblivion: even the classification of the chemical elements – paradigmatic natural kinds by virtually every philosopher's lights – inclines us towards focussing on explanatory and predictive success rather than sameness of constitution.

With these clarificatory remarks in place, we now proceed to examine three approaches to natural kinds. The first two – what we shall call the 'stipulative account' (§3) and the 'Kripkean account' (§4) – fall roughly into the category of the 'constitutional' approach described above, while the 'homeostatic property cluster' (HPC) account (§5) explicitly focuses on explanatory and predictive success. Our aim is to argue (briefly) that the stipulative and Kripkean accounts are problematic; that the HPC account therefore deserves to be taken seriously as an alternative account of natural kinds; and that in principle psychiatric kinds might well fit the HPC account.

3. Conventional vs. natural kinds: the stipulative account

Concerns about the distinction between realism and conventionalism about classification go back to at least John Stuart Mill, and probably even further. The debate is often framed in terms of natural kinds, and, as Mill noted, "a natural classification is grounded on real kinds, its groups are certainly not conventional; it is

perfectly true that they do not depend upon an arbitrary choice" (Mill 1869: IV, Ch. II). For Mill, then, a 'natural' classification carves nature at its joints – that is, *bona fide* divisions in nature. But how are we to determine which are the 'real kinds' and which are merely 'conventional'?

We shall approach this issue by outlining the features that are generally thought to characterise conventional kinds, as a way of highlighting the sorts of features that *non*-conventional – that is, natural – kinds might be thought to have. First, what we might call 'relativity'. The category *tasty food* is a classification that is relative to the gustatory preferences of the classifier: whether or not chicken vindaloo falls into that category varies from person to person, and so is clearly not 'objective' in at least one sense of the word.

Second, vagueness: there is no sharp boundary in nature between tall and non-tall people – just particular facts about the individual heights of individual people. We rather roughly divide them into the tall and the non-tall, but some people are borderline-tall: there is no fact of the matter about whether or not they are tall.

Third, arbitrariness: we might define, say, a *ravcat* as something that is a raven or a cat. Even if the kinds *raven* and *cat* are themselves natural, *ravcat*, arguably, is not: it is mere linguistic stipulation that, as it were, brings this new kind into being. A second way arbitrariness can arise is as a response to vagueness: we might 'precisify' a vague concept, such as 'tall person', by stipulating that a person is tall if and only if they are, say, at least six foot seven. Then the category *tall person* would not be vague; nonetheless, it would be arbitrary. Nothing in nature motivates us to draw the line at six foot seven as opposed to, say, six foot six-and-a-half.

Fourth, and finally, as we have already seen, it has often been assumed that the, or a, distinctive mark of natural kinds is that their members share a 'real essence', in the sense that there is some underlying feature, present in all members of the kind, that explains the observable characteristics and behaviour of those members. For example, the molecular composition and structure of water (a paradigmatic natural kind as far as most philosophers are concerned) explains why it is liquid at room temperature, has a boiling point of 100°C in normal atmospheric conditions, and so on.

The above list is supposed to be merely indicative; we do not take ourselves to have provided a thorough and complete analysis of the notion of a conventional kind. Nonetheless, the list does suggest one way in which one might attempt to characterise *natural* kinds, which is to *stipulate* that natural kinds are those kinds that lack the features that conventional kinds (according to the above characterisation) lack; we shall call this the 'stipulative account' of natural kinds. Thus we might stipulate that natural kinds are kinds that (i) are not relative to human interests, (ii) are not vague, (iii) are not stipulatively defined, and (iv) share an underlying 'real essence'.

This is (one part of) the strategy adopted by Brian Ellis (2001, 19-23), who stipulates that conditions (i)-(iv) must be met by any *bona fide* natural kind (along with some other requirements that we shall ignore).

The major problem with the stipulative account, we think, is that it places conditions

on natural kinds that are far too onerous. While the listed conditions are *prima facie* plausible, in that they clearly rule manifestly conventional kinds (such as *tasty food*, *tall person* and *ravcat*) as non-natural, they are in fact in serious tension with a plausible story about what classification according to natural kinds is supposed to achieve, namely (to put it roughly) predictive and explanatory success.

To see this, consider the case of biological species. There are two main 'species concepts' in biology. The first is a genetic concept: a species is defined in terms of some sort of genetic similarity that holds between all and only members of a given species. The second is the cladistic concept: to put it very roughly, species are carved out according to 'nodes' on the tree of life, so that the point at which the tree branches is the point at which a new species comes into existence. On this view, the essence of a species is a historical matter: two animals that occupy two distinct branches on the evolutionary tree are members of two distinct species, no matter how genetically similar they are to one another.

The stipulative account rules out species as natural kinds on either view of species essences. Genetically-defined species are not natural kinds because there simply is, as a matter of empirical fact, no single, identifiable genetic trait or collection of traits that all and only members of a given species share. This does not undermine the genetic conception of species – one might still be able to provide a genetic characterisation of a given species in terms of, say, distinct, individually non-necessary but sufficient sets of genetic traits, so that two animals (or whatever) might each satisfy a different set of conditions and thus qualify as members of the same species despite having no one relevant trait or set of traits in common. But it does undermine the claim of genetically-defined species to be natural kinds according to the stipulative account, because species, so-defined, violate the condition that members of a natural kind must share a common nature. Cladistically-defined species are ruled out on the grounds that essences must be *intrinsic*. That a given animal is descended from animals that occupy a particular position on the evolutionary tree is not an intrinsic feature of it, and so cannot constitute the essence of any natural kind to which the animal might belong. Put simply, historical facts cannot constitute natural-kind essences, according to the stipulative account.

Of course, the obvious response to make on behalf of the stipulative account would be simply to deny that biological kinds are genuine natural kinds. After all, one might try to claim, once we understand the facts that determine membership of these kinds, those facts clearly do not mark out any natural boundaries. But this in turn undermines the thought that natural-kind classification is something that scientific enquiry *should* aim at. For example, we know enough about evolutionary biology to be reasonably certain that *no* species concept will meet the conditions laid down by the stipulative account. Where does that leave evolutionary biology? The scientific credentials of evolutionary biology are not in doubt: it is undeniable that the concepts *species*, *genus* and *family* are highly predictively and explanatorily useful, as are the more specific concepts *cat*, *amoeba*, *coyote*, and so on. Indeed, Ellis himself concedes that we can "reason about [biological species] as if they were strict Aristotelian natural kinds" (Ellis 2001, 170). Moreover, these concepts clearly carve nature at its joints

in *some*, more permissive, sense: the division of the animal kingdom into cats, dogs, rodents, and so on is certainly not arbitrary or in any sense merely a function of our interests.

The moral for the philosophical issue surrounding the 'reality' or 'objectivity' of psychiatric kinds is clear. If we adopt the stipulative account of natural kinds, then psychiatric kinds are never going to count as natural kinds, since the only kinds that are likely to count as natural will be those found in fundamental physics and (some of) those found in chemistry. Thus even if, say, a genetic or neurological basis was found for psychiatric kinds, this would not be sufficient to establish their status as natural kinds, since genetic and neurological classifications themselves will fail to 'carve nature at its joints' in the sense required by the stipulative account.

This in no way impugns the scientific credentials of psychiatry, however – just as it does not impugn the scientific credentials of evolutionary biology or genetics – and nor would it entail that psychiatric kinds are somehow predictively or explanatorily deficient (again, as with biology and genetics). We can be pretty sure that, given the stipulative account, *no* natural-kind classification will deliver any remotely viable psychiatric predictions or explanations, since physics and chemistry are not the places to look for explanations of mental phenomena generally.

We suggest, therefore, that the stipulative account is not a good answer to the question, 'what is it for something to be a natural kind?'. We have no particular grounds for endorsing an account of natural kinds that renders biological species, say, non-natural even though we can 'reason about them as if' they *were* natural kinds. After all, consider the motivation for the stipulative account. We approached this question by listing some features that conventional kinds have, so that natural kinds are then defined in terms of the absence of those features (relativity, vagueness, and so on). But, as it turns out, this approach is simply too restrictive: *some* of the features had by some conventional kinds are also shared by some (intuitively) natural kinds. So there is no reason not to look for a less restrictive conception of 'natural kind' – one that rules out *tasty food*, *tall person* and *ravcat*, while ruling in *raven*, *cat* and *tiger*.

4. The Kripkean account

The second kind of approach to natural kinds – which we shall call the 'Kripkean account' – takes its lead from a Kripkean story about the semantics of so-called 'natural kind terms'. According to Kripke (1980) there is a special class of general terms, the 'natural kind terms', that are remarkably similar to proper names. As is well known, according to Kripke proper names are best understood as having denotation (or 'reference') but no connotation ('sense'). Thus, for example, the name 'Sigmund Freud' serves merely to pick out a particular man, Freud, rather than being synonymous with some description ('the inventor of psychoanalysis', say). Moreover, names like 'Sigmund Freud' are supposed to be *rigid designators*: they refer to the same object (in this case, Freud) in all possible worlds. According to Kripke, natural kind terms (such as 'water', 'gold' and 'tiger') are also supposed to have denotation

but no connotation, and designate the same kind in all possible worlds.

The interesting consequence of the Kripkean account is that identity sentences containing co-referring proper names, such as 'Charles Dodgson is Lewis Carroll' that are knowable only *a posteriori* – it takes some empirical investigation to discover that Dodgson is Carroll – turn out to be necessary, given the rigidity of proper names plus the principle that every object is necessarily self-identical (the necessity of identity). Similarly, so-called 'theoretical identities', such as 'gold is the element with atomic number 79', are necessarily true – gold is the element with atomic number 79 in all possible worlds in which it exists – but knowable only *a posteriori*. Scientific investigation *discovers* "the nature, and thus the essence" of gold (Kripke 1980, 138); it does not *define* 'gold' to mean 'the element with atomic number 79'.

The Kripkean account is in principle more permissive than the stipulative account, in that, just by itself, it does not require natural-kind essences to be a matter of shared intrinsic features. To see this, consider the analogy with the 'necessity of origin thesis'. Kripke himself holds that people have the parents they actually have essentially: nobody who failed to be the biological offspring of Randolph Churchill and Jennie Jerome *could* be Winston Churchill. So, presumably, it is supposed to be metaphysically necessary, but knowable only *a posteriori*, that Churchill is the son of Randolph and Jennie. Similarly, then, on the Kripkean account biological kinds understood cladistically can in principle count as natural kinds, so that their essences are constituted by historical, evolutionary facts.

The Kripkean account thus looks *prima facie* more promising as an account of natural kindhood, in that it might in principle render psychiatric kinds natural. Several authors (e.g. Papineau 1994) have suggested that psychiatric classification might be possible in terms of the notion of 'maladaptive traits' or 'biological disfunction'. If this did turn out to be possible, then psychiatric kinds might turn out to have evolutionary, historical essences, just as (cladistically-conceived) biological species do; and so they could, by Kripkean lights, count as natural kinds.

We believe, however, that there are no good reasons to accept the Kripkean account, because we see no reason to suppose that the *semantic* category of 'natural kind terms' will serve to mark out a *metaphysically* distinctive category, viz, the natural kinds. The Kripkean account, recall, starts from a linguistic thesis: a thesis about the conditions under which certain *linguistic* phenomena, namely some 'general terms', have a certain semantic feature, namely being such that they rigidly designate a 'kind', the conditions for membership of which are discovered by empirical investigation rather than featuring in a stipulative *definition* of the kind (that is, rather than constituting the *meaning* of the kind term). Kripke himself calls the kinds that are named by terms having this feature the 'natural kinds'. What he does not do, and what to our knowledge nobody has done, is provide any argument for the claim that such kinds carve nature at its joints, in any metaphysical sense.

There are good reasons to be sceptical about the prospects of such an argument. First, the Kripkean account appears to be too permissive. Many philosophers of language now take many more general terms to be rigid designators than are, intuitively, natural kind terms.

For example, Genoveva Martí and Jose Martínez-Fernández argue that all 'semantically simple' general terms, such as 'bachelor', rigidly designate kinds (Martí and Martínez-Fernández 2010). If that is right, then rigid designation cannot be the mark of a natural kind term.

Second, the Kripkean semantic story appears to rule that only pretheoretical, vernacular terms can be natural kind terms: 'water', 'gold', 'tiger', and so on. This is because the meta-semantic underpinning of the Kripkean account – the idea that there is an 'initial baptism' of the kind – presupposes that the kind is named in the absence of knowledge of its essence. But, clearly, there are not nearly enough vernacular terms to pick out *all* the natural kinds. For example, ununbium (assuming it really has been created in the lab) – the element with atomic number 112 – is clearly a natural kind if gold is. But the name for this kind – 'ununbium' – is stipulatively defined according to rules laid down by the International Union of Pure and Applied Chemistry: if the element allegedly created by scientists turned out to be the element with atomic number 113, IUPAC would rule that ununbium had not, in fact, been created, and not that ununbium turns out to have a different essence to the one we thought it had. In other words, 'ununbium is the element with atomic number 112' is knowable *a priori*; there is *no* necessary truth knowable only *a posteriori* of the form '… is the element with atomic number 112'. (See Beebee and Sabbarton-Leary 2010 for more on this.)

The moral is that there is no reason to think that the availability of theoretical identity statements that are metaphysically necessary but knowable *a posteriori* is a distinctive mark of the presence of a natural kind. This is crucial for the issue of psychiatric kinds because one argument that might be given for the failure of such kinds to count as natural is the fact that psychiatric kinds are stipulatively defined, and hence their 'essence' is knowable *a priori*. According to Brian Ellis, this does indeed, just by itself, serve to rule out the possibility that such kinds are natural kinds, since the ground of an analytic truth is merely a "conventionally established criterion for including something in some linguistically defined class" (Ellis 2001, 235), whereas the ground for the necessity of a true theoretical identity concerning natural kinds is "radically non-linguistic and objective" (Ellis 2001, 36). We hope to have shown that this charge is, at least at present, unwarranted – in the case of psychiatric kinds no less than in the case of ununbium.

This leaves us back where we started: in search of a metaphysical account of the nature of natural kinds. We have already considered the stipulative account; we shall now examine a rather more permissive account.

5. A more permissive conceptions of natural kinds: the 'homeostatic property cluster' account

Let us return to the question, what is classification in terms of natural kinds supposed to *achieve*? As we saw in §3, Ellis claims that we can treat biological kinds *as though* they are natural kinds, in that biological classifications are explanatorily and predictively successful, even though (on his view) they are not natural kinds because

they fail to meet the criteria laid down by the stipulative account. But we might take another approach here: as we said in §2, we might *start* from the thought that the natural kinds are just those classifications that are explanatorily and predictively successful. After all, Ellis himself says that we classify objects into kinds on the basis of a shared essence, and do so because "essences are the postulated intrinsic sources of the manifest properties and behaviour[s]" of members of the kind, and hence ground the inductive and explanatory success of science (Ellis 2001, 92). If this is supposed to be the *point* of natural-kind classification – which would seem to be obligatory if we are to think of the sciences as *aiming* at uncovering the natural kinds – then we are much more likely to come up with an adequate conception of natural kinds if we take explanatory and predictive success as the distinguishing feature of natural-kind classification, rather than a feature that merely happens to be correlated (and imperfectly so on Ellis's view, as the biological case illustrates) with some *other* distinguishing feature.

While there is more than one account that takes this general approach on the market (see in particular Dupré 1993), the one we shall focus on here is Richard Boyd's 'homeostatic property cluster' or 'HPC' account (see for example Boyd 1991, 2010). For Boyd, the "fundamental question which the theory of natural kinds addresses is this: 'How do classificatory practices and their linguistic manifestations help to underwrite the reliability of scientific (and everyday) inductive/explanatory practices?' " (2010, 215). His answer involves the idea of a 'fit' between the inferential practices (involving use of the relevant kind terms) of a given discipline on the one hand, and causal structures in the world on the other. So a *natural* kind cannot be a kind, membership of which is *merely* a matter of arbitrary stipulation or linguistic convention; natural-kind classification is marked out by 'fit' with causal structures, so that predictive and explanatory success reflects those causal structures. In other words – to put it rather simplistically – natural-kind classification will not simply *in fact* happen to be predictively and explanatorily successful; its success is explained by the fact that it latches on to real causal mechanisms.

In some cases of natural kinds, such a fit is achieved by appeal to intrinsic underlying features shared by all and only members of the kind; this is so in the case of, say, the chemical elements. But such a unified underlying intrinsic nature is not *required* for a kind to count as 'natural', on Boyd's view. In particular, many natural kinds (including biological and, Boyd claims, at least some social kinds) will be characterised by 'clusters' of properties, where the 'clustering' is maintained by what he calls 'homeostatic' causal mechanisms. The full cluster of properties need not be shared by all members of the kind; nor need there be *one* single mechanism that causes the presence of these features (just as the maintenance of a stable blood pressure – a homeostatic mechanism in the strict biological sense of 'homeostatic' – is effected by a large range of cardiovascular, neuromuscular and hormonal mechanisms). So, for example, features such as having four legs, two eyes, a tail and fur are typical features of cats, but not universal ones. But there are causal mechanisms – genetic transmission, environmental pressures, and so on – that are responsible for the (usual) possession of all of these traits in cats. It is the 'homeostatic' nature of these mechanisms that allows us to use the category *cat* both to explain and to predict the features and behaviour of

particular cats, although, again, not with perfect reliability.

The precise semantic and metaphysical consequences of Boyd's view are too complex to deal with here; instead, we shall simply summarise the most important features for the purposes of this paper. First, the distinction between natural and non-natural kinds resides in the fact that the former, but not the latter, are determined by the 'fit' of our classificatory, inferential and explanatory practices with real causal mechanisms. Boyd says that "the unity of the property-cluster which defines [natural property-cluster kinds] is *causal rather than conceptual*" (1991, 141), and often says that the definitions of kind terms are to be discovered *a posteriori* rather than *a priori*. However, we take it that this enshrines no commitment to a Kripkean semantics. Even if it is analytically true that, say, ununbium is the element with atomic number 112, the classification of elements by atomic number is itself motivated by the empirical findings of chemistry, and not merely the arbitrary stipulation of armchair chemists. Indeed, classification by atomic number has persisted precisely *because* of the explanatory and predictive success of so doing, which in turn is due to the fact that the causal mechanisms at work in particular members of a given element kind deliver similar features and behaviour.

Second, the account is, as should be obvious, considerably more permissive than traditional accounts of natural kinds (and in particular the stipulative account): there is no requirement that members of a natural kind share any underlying intrinsic feature, or that natural kinds have sharp boundaries, or (as we argued above) that they cannot be stipulatively defined. Indeed, Boyd explicitly claims that the account can accommodate *social* kinds. Here he takes issue with Ian Hacking's claim that the defining properties of a natural kind are 'natural rather than social' (Boyd 1991, 127; see also Hacking 1991). On Boyd's view, as we have seen, what is distinctive about natural kinds is that they are underpinned by genuine causal mechanisms; and nothing in that requirement rules out social kinds in principle (though of course it may rule out *some* social kinds). Paul Griffiths notes that *money* plausibly counts as a natural kind on the HPC account:

> Money … is a key node in many economic theories. The lawlike generalizations about money, such as those connecting money supply to inflation or interest rates, hold true in an economy because of a social convention treating some class of objects as a means of exchange and because agents in that economy try to maximize their utility. Neither of these circumstances is linked to any intrinsic property of the currency units. (Griffiths 1999, 218)

Of course, the permissiveness of Boyd's account might lead one to suspect that Boydian natural kinds do not, after all, cut nature at its joints. Boyd says:

> Kinds useful for induction or explanation must always 'cut the world at its joints' in this sense: successful induction and explanation always require that we accommodate our categories to the causal structure of the world (Boyd 1991, 139).

To put things slightly differently: the distinction we drew at the beginning of §3 was a distinction between natural and 'merely conventional' kinds. Boyd holds that the requirement that natural-kind categories are 'accommodated' to (or 'fit') the causal structure of the world just *is* the requirement that such categories carves nature at its joints, at least in one sense of 'carving nature at its joints'. 'Cat' and 'money' are not merely conventional-kind terms, since they fit the causal structure of the world, and notwithstanding the fact that, in the case of money, that causal structure is itself a product of social practices. To respond that such a distinction does not *really* adequately capture the thought that natural-kind classifications should carve nature at its joints seems to us to rest on what we called in §2 a 'constitutional' conception of natural kinds – a conception whose viability we have already called into question and which in any case is distinctly optional.

Of course, the issue of whether or not psychiatric kinds might count as natural kinds on a Boydian view is still open: for all that has been said, perhaps psychiatric kinds fall on the 'conventional' rather than the 'natural' side of Boyd's divide. It seems clear to us, however, that at least some psychiatric kinds could in principle be natural kinds in Boyd's sense.

The most obvious feature of psychiatric classification (as determined by the *Diagnostic and Statistical Manual of Mental Disorders* (DSM)) is that it is 'prototypical': there is generally no one set of necessary and sufficient conditions for the diagnosis of the disorder, but rather a list of symptoms, of which the patient must display sufficiently many. As should be clear, this feature of psychiatric kinds is no bar to their counting as natural kinds according to HPC; indeed, this is a feature that psychiatric kinds share (according to Boyd) with biological kinds.

This being so, the key remaining issue is whether psychiatric kinds generate inferential practices – roughly, explanation and prediction – that have the required 'fit' with causal structure. And the prospects here, at least in principle, are promising (though of course a more detailed account is needed). That psychiatric kinds play, or at least can play, a *bona fide* role in explanation and prediction is surely beyond dispute: knowing that someone has Tourette's or is bipolar allows one to draw a good number of inferences concerning likely behaviour, what forms of treatment are likely to succeed, and *how* likely it is that a given treatment will succeed. Of course, such predictions are not infallible; they are not like the predictions of fundamental physics or of chemistry. But the same is true of biology: that something is a cat does not *guarantee* that it has four legs or fur or a tail, or that it will eat certain kinds of food and not others, and so on. And it is at least *prima facie* reasonable to suppose that such inferences latch on, or in principle *can* latch on, to real causal structures. After all, presumably there is a *reason* why the classic symptoms of, say, schizophrenia or Tourette's tend to cluster together; these categories are not defined by arbitrarily-selected collections of properties that merely happen, in so-far-observed cases, to co-occur.

Crucially, however, this is not to say that there is *one* reason why the relevant properties cluster together across all cases. There might be a variety of causal bases that deliver the relevant symptoms (neurological, developmental, genetic, chemical, etc.), such that

these causal bases themselves vary from one (say) schizophrenic person to the next. This would not be in tension with the characterization of schizophrenia as a natural kind, so long as differences in causal basis are largely irrelevant to predictive and explanatory success. For example, were it to turn out that two distinct neurological bases (call them N1 and N2) for schizophrenia were discovered, such that it turned out to be much more explanatorily and predictively useful to classify according to N1 and N2 separately rather than according to the current prototypical classification according to mental and behavioural features, then that would be a good reason to reject schizophrenia as a natural kind and replace it with the relevant neurological classification. On the other hand, if it turned out that whether someone possessed neurological basis N1 or N2 made no significant difference to, say, behaviour, prognosis or treatment, there would be no grounds for reclassifying, since on the HPC view intrinsic sameness of underlying causal basis is not a requirement for natural kinds.

The above is of course only a brief sketch. However, the moral, we think, is clear enough: given the HPC account, there are, at least *prima facie*, no principled grounds for denying that psychiatric kinds can be natural kinds.

6. An objection disarmed

There is one immediate worry that one might have about the claim that psychiatric kinds can be natural kinds, which we briefly mentioned in §2. While the HPC account is (and is designed to be) a permissive account – and in particular it is designed to accommodate the possibility that *some* social kinds are natural kinds – one might claim that the notion of a psychiatric *disorder* manifestly falls on the 'socially constructed' side of the line, and so psychiatric kinds cannot count as natural kinds after all. This worry is raised by, for example, Szasz's (1961) objection to the concept of 'mental illness', and has pervaded discussion of psychiatric kinds ever since.

Szasz's basic beef with the notion of 'mental illness', as we understand it, is that it miscategorises what he calls 'problems with living' – problems surrounding human relationships generally – as illnesses, susceptible to medical treatment. On Szasz's view, treating problems with living as though they are illnesses deflects attention away from where it is properly located: to put it simply, it mistakes what are in fact symptoms of social disharmony for treatable maladies, and thus 'treats' the symptoms instead of the cause. Just as "fighting the battle of stomach acid and chronic fatigue instead of facing up to marital conflict" is an example of "waging battles on false fronts" (1961, 118), so is treating a psychiatric "illness" by medical means a failure to address the underlying causes of that "condition". Many other authors have noted that what counts as a 'disorder' is liable to significant revision in the light of changes in public morality (homosexuality being the most obvious case), and that the extent to which a given condition is detrimental to a person's wellbeing depends to a great extent on the nature of the social environment (including norms of behaviour) in which they live. All this strongly suggests that the kind *psychiatric disorder* does not itself constitute a natural kind, and so the subclassification into *kinds* of disorder will inherit this lack of naturalness.

As should be clear from the discussion of §2, however, we believe that this problem is not as thorny as it is sometimes taken to be. We take our cue from Szasz himself, who says: "While I have argued that mental illnesses do not exist, I obviously did not imply that the social and psychological occurrences to which this label is currently being attached also do not exist" (1961, 117). It is thus consistent with Szasz's own position to claim that certain *conditions* – Tourette's, say, or schizophrenia – exist, and indeed are natural kinds, even if their *status as disorders* is itself socially constructed in a way that renders the kind *psychiatric disorder* non-natural.

Admittedly, such a distinction between (potentially natural-kind) classification of particular psychiatric conditions and (non-natural-kind) classification of those conditions as *disorders* fits badly with existing nomenclature on several fronts. Many such conditions have 'disorder' in their names (bipolar disorder, for example); and indeed the very label '*psychiatric* condition' is itself arguably lacking in neutrality. Etymologically, the '-iatry' suffix implies healing, and of course psychiatry itself is by its nature concerned with *both* classification *and* treatment: the very *point* of psychiatric classification is that it aids treatment. Even so, the conceptual distinction between the classification of a given psychiatric condition and the classification of that condition *as a disorder* (or, if you like, as a genuinely *psychiatric* condition) can (at least in principle) still be made. (Homosexuality, for example, has some claim to count as a natural kind on a permissive conception of natural kinds; but if it is a natural kind, it was a natural kind even when it was (mis)classified as a psychiatric disorder.) To put the point another way, whether or not some category (schizophrenia or homosexuality or whatever) is a natural kind is one question; whether or not it ought to be listed in the *DSM* is separate question.

It might reasonably be objected at this point that many psychiatric disorders are *defined* in term of their status as disorders; for example, criterion C of a major depressive episode states: "The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning" (American Psychiatric Association 2000, 356), and it could be argued that whether or not such 'impairment' exists is too contingent a feature of particular social or occupational situations for *major depressive episode* to constitute a genuine natural kind.

Our response to this objection is threefold. First, our intention in this paper is not to argue for the claim that *all* psychiatric kinds are natural kinds. It is rather the significantly weaker claim that there is no *principled* reason why at least *some* psychiatric kinds cannot be natural kinds. Perhaps some psychiatric kinds will turn out *not* to be natural kinds because there are no sensible classification criteria that do not appeal to social or occupational impairment.

Second, reference to 'significant distress or impairment' is itself generally present in order ensure that there are not too many 'false positives'; and a major reason why lack of false positives is deemed desirable is that (to put it rather flippantly) it wouldn't do for official psychiatric classification to result in a large proportion of the population counting as suffering from a psychiatric disorder. (Most people suffering less significant distress or impairment would be much better off changing

jobs or thinking hard about their relationships than consulting a psychiatrist.) But in our view this merely highlights the fact that the need for the 'significant distress or impairment' condition derives from the likely effects of classifying someone as having a psychiatric disorder – as opposed to classifying them as merely having, say, a 'depressive episode'. And this is entirely consistent with the conceptual distinction we are pressing. In principle, one might be able to separate out a distinctive natural kind ('depressive episode', say) without appealing to significant distress or impairment; the reason why this is not, in fact, done is that it conflicts with the need to classify psychiatric *disorders*. If we're doing *that*, we need to raise the bar in order to rule out false positives. But the false positives are not false positives for some natural kind ('depressive episode', say); rather they are false positives for some non-natural *disorder* kind ('major depressive episode'). In other words, there could well be a natural kind in the offing; it's just that for purposes to do with who should be regarded as a legitimate candidate for psychiatric treatment, members of a (non-natural) subset of that natural kind are singled out as sufferers of a disorder.

Third, and finally, all of the above grants that the kind *disorder* is not itself a natural kind; and this assumption can be questioned. If it is found to be unwarranted, the problem that we have sought to address in this section goes away. And there are some grounds for optimism here. After all, as we have seen, Boyd himself takes HPC to be consistent with the existence of *social* natural kinds; and, for all that has been said so far, *psychiatric disorder* could turn out to be one such kind. In particular, the fact that some kind's counting as a psychiatric disorder might depend on highly contingent facts about the social environment does not automatically render *psychiatric disorder* a non-natural kind if at least *some* social kinds count as natural, since of course *any* social kind's explanatory and predictive utility – and thus its status as a natural kind – will depend upon facts about the social environment. Even distinctively normative features of that environment (e.g. what is generally deemed to be socially acceptable behaviour) need not rule out the relevant kind from counting as natural. Boyd himself suggests that racial kinds might be "natural kinds in the social sciences that study stratification, poverty and political oppression" (2010, 222) (while denying that this implies that race must also be seen as a natural kind in any other discipline, such as biology). And yet the 'natural' status of racial kinds within the social sciences depends on unpleasant but nonetheless real facts about social norms (in particular, promulgation and tolerance of racist behaviour). Similarly, then, the (alleged) fact that social norms play a role in categorisation in terms of psychiatric disorders need not, in principle, be a bar to such a categorisation being a natural-kind classification. Nor need such categorisation involve endorsing those norms. In principle one can categorise a psychiatric kind as a disorder while wishing that the social norms that are responsible for the social and occupational impairment that membership of that kind brings in its train were otherwise, just as one can investigate the relationship between race and poverty or political oppression without endorsing the norms that are responsible for that relationship.

REFERENCES

American Psychiatric Association. 2000. *Diagnostic and Statistical Manual of Mental Disorders, 4th Edn. (Text Revision)*.

Beebee, H. and Sabbarton-Leary, N. 2010. On the abuse of the necessary *a posteriori*. In *The Semantics and Metaphysics of Natural Kinds,* ed. H. Beebee and N. Sabbarton-Leary, 159-78. New York: Routledge.

Bolton, D. 2008. *What is Mental Disorder?* Oxford: Oxford University Press.

Boyd, R. 1991. Realism, anti-foundationalism and the enthusiasm for natural kinds, *Philosophical Studies* 61: 127–48.

Boyd, R. 2010. Realism, natural kinds, and philosophical methods. In *The Semantics and Metaphysics of Natural Kinds,* ed. H. Beebee and N. Sabbarton-Leary, 212-34. New York: Routledge.

Cooper, R. 2005. *Classifying Madness: A Philosophical Examination of the Diagnostic and Statistical Manual of Mental Disorders.* Dordrecht: Springer.

Dupré, J. 1993. *The Disorder of Things*. Cambridge, MA: Harvard University Press.

Ellis, B. 2001. *Scientific Essentialism*. Cambridge: Cambridge University Press.

Griffiths, P. 1999. Squaring the circle: natural kinds with historical essences. In *Species: New Interdisciplinary Essays,* ed. R. Wilson, 209-28. Cambridge, MA: MIT Press.

Hacking, I. 1991. A tradition of natural kinds. *Philosophical Studies* 61: 109-26.

Kripke, S. 1980. *Naming and Necessity.* Oxford: Blackwell.

Martí, G. and Martínez-Fernández, J. 2010. General terms as designators: a defence of The View. In *The Semantics and Metaphysics of Natural Kinds,* ed. H. Beebee and N. Sabbarton-Leary, 46-63. New York: Routledge.

McCrae, R. R. 1994. 'A reformulation of axis II: Personality and personalit related problems' in *Personality disorders and the five-factor model of personality*, ed. Costa, T. & Widiger, T. A., 303-10. Washington, DC: American Psychological Association.

McGinn, C. 1978. Mental states, natural kinds, and psychophysical laws. *Proceedings of the Aristotelian Society*, Supp. Vol., 52: 195-220.

Mill, J. S. 1869. *A System of Logic.* New York: Harper and Brothers.

Papineau, D. 1994. Mental disorder, illness and biological disfunction. In *Philosophy, Psychology and Psychiatry*, ed. A. Philips-Griffiths, 73-82. Cambridge: Cambridge University Press.

Szasz, T. 1961. The myth of mental illness. *American Psychologist* 15: 113-8.

Zachar, P. 2000. Psychiatric disorders are not natural kinds. *Philosophy, Psychiatry, and Psychology* 7: 167-82.

Received: March 14, 2010
Accepted: April 13, 2010

Philosophy Department
University of Birmingham
Birmingham B15 2TT UK
h.beebee@bham.ac.uk
nigelleary@hotmail.com